用小波包变换产生音频数据索引的方法

李 应,侯义斌
(西安交通大学电脑与信息技术研究所,陕西西安 710049)

摘 要: 针对音频多媒体数据库中基于例子和基于内容的查询,本文提出一种产生音频数据索引的方法.这里,我们首先讨论了小波包分解的过程和最好基及代价函数的选择方法.其次,对现有的用小波变换产生音频数据索引的二个方法进行比较,并提出基于小波包最好基变换产生音频数据索引的方法.再次,我们提出用音频数据的小波包最好基变换系数的部分最高值的能量作为音频数据索引.最后,我们把这种方法与直接采用小波变换产生索引的方法相比较.实验结果表明这种新方法具有较高和较稳定的检索精度.

关键词: 音频数据; 数据检索; 小波变换; 小波包; 最好基选择
中图分类号: TP311.134.3, TP912.3 **文献标识码**: A **文章编号**: 0372-2112 (2003) 04-0593-04

Method of Using the Wavelet Packet Transform Derivation Audio Data Index

LI Ying, HOU Yi-bing
(*Computer and Information Technology Insititute Xi'an Jiaotong University, Xi'an, Shannxi* 710049, *China*)

Abstract: A method for deriving the audio data index is proposed in this paper, which aim at query-by-example and query-by-content in audio multimedia databases. Here, first we discuss the process of wavelet packet decomposition and the method for choosing the best base and cost function. Second we prepare two schemes, which has been used, that deal with deriving the index of wavelet transform. And we propose the method to derive the audio data index that it is based on the best base transform in the wavelet packet. In the third, we take the part of energy for the audio data index which come from the highest value coefficients of the best base transform in the wavelet packet. In the last, we take our method comparing with the method that derived the index directly form the wavelet transform. The experiment shows that there are higher and robust for retrieving precision in the new method.

Key words: audio data; data retrieval; wavelet transform; wavelet packet; best base choose

1 引言

与图像和视频等其它多媒体数据一样,音频数据需要基于内容的索引和近似搜索.而支持基于内容和例子回取的音频多媒体数据库,应当能够自动分析音频数据而生音频数据的索引.文献[1]描述了基于内容的音频的分类、搜索和回取系统.它主要用统计技术来分类、分析声音,并得出由均值、变量、自相关和延续等组成的特征向量,用特征向量作为基于内容的分类和声音回取的唯一信息.在文献[2]中,描述了通过例子查询的搜索和回取系统,但它只支持音乐数据和查询一些较规则的震动声音.它们都存在计算量大而不精确的问题.对于可以管理任何种类的音频数据的索引的产生方法,文献[3]提出了通过短时 Fourier 变换和小波变换产生索引的方法.采用小波变换产生索引比基于信号统计和基于短时 Fourier 变换产生索引的方法具有更好的回取精度.此外,与小波变换相关的小波包变换也已开始用于信号处理、语音编码和压缩[4~7],但却很少在音频多媒体数据库使用小波包来产生索引.

本文在文献[3]中采用的基于小波变换产生音频数据索引的基础上,提出用小波变换产生音频数据索引的方法.这种方法可以根据不同类型的音频数据,通过小波包变换及代价函数选择最好基.用不同的最好基对音频数据进行变换,并根据变换系数分布,选择一定比例的系数最大值的能量作为索引系数.这种方法使得基于例子的音频数据查询的回取精度得以较大提高.

2 小波包分析

2.1 小波包的空间分解

音频信号是非平稳信号,由于小波变换具有多分辨率分析的特点,它比短时 Fourier 变换更适合用来抽取音频数据信号的特征,而对于不同类型的音频信号,它们的频谱差别较大,因此我们采用小波包来分析各种类型的音频信号.

对于小波多分辨率分析,可以按照不同的尺度因子 j 把 Hilbert 空间 $L^2(R)$ 分解为所有小波子空间 $W_j(j \in Z)$ 的正交和,即 $L^2(R) = \bigoplus_{j\in Z} W_j$. 令 $U_j^0 = V_j$, $U_j^1 = W_j$, $j \in Z$,则 Hilbert 的空间的正交分解 $V_{j+1} = V_j \oplus W_j$ 可表示为

收稿日期:2001-07-16;修回日期:2002-07-08

$$U_{j+1}^{0} = U_{j}^{0} \oplus U_{j}^{1} \tag{1}$$

令 $k = 1,2,\cdots, j; j = 1,2,\cdots$,并对式(1)作迭代分解,则有

$$W_j = U_j^1 = U_{j-1}^2 \oplus U_{j-1}^3$$

$$U_{j-1}^2 = U_{j-2}^4 \oplus U_{j-2}^5, U_{j-2}^5 = U_{j-2}^6 \oplus U_{j-2}^7,$$

$$\cdots$$

对 W_j 进行再分解,得

$$\left.\begin{aligned} &W_j = U_{j-1}^2 \oplus U_{j-1}^3 \\ &W_j = U_{j-2}^4 \oplus U_{j-2}^5 \otimes U_{j-2}^6 \oplus U_{j-2}^7 \\ &\cdots \\ &W_j = U_{j-k}^{2^k} \oplus U_{j-k}^{2^k+1} \oplus \cdots \oplus U_{j-k}^{2^{k+1}-1} \\ &\cdots \\ &W_j = U_0^{2^j} \oplus U_0^{2^j+1} \oplus \cdots \oplus U_0^{2^{j+1}-1} \end{aligned}\right\} \tag{2}$$

即 W_j 空间分解的子空间序列可写作 $U_{j-1}^{2^l+m}$,其中 $m = 0, 1, \cdots, 2^l - 1; l = 1,2,\cdots, j; j = 1,2,\cdots$. 对于每个 $j = 1,2,\cdots$,由式(2)可以得出

$$L^2(R) = \bigoplus_{j\in Z} W_j = \cdots \oplus W_{-1} \oplus W_0 \oplus U_0^2 \oplus U_0^3 \cdots \tag{3}$$

若令 $n = 2^l + m$,小波包记为

$$\Psi_{j,k,n}(t) = 2^{-j/2}\Psi_n(2^{-j}t - k)$$

其中 $\Psi_n(t) = 2^{l/2}u_{2^l+m}(2^l t)$. 这时,

$$\{u_{j,k}(t), u_n(t-k) \mid j = \cdots, -1, 0; n = 2,3,\cdots, \text{且 } k \in Z\}$$

是 $L^2(R)$ 的一个正交基.

从式(2)和(3)可知,小波包可以对 W_j 进一步分解. 它克服了小波多分辨率分解中时间分辨率高时,频率分辨率低的缺陷,因此具有更好的视频特性.

2.2 最好基的选择

为了有效地抽取音频信号的特征,对于不同类型的音频数据的变换,应当使用不同的最好基. 选择最好基,首先要产生代价函数. 常用方法的是把 Shannon 熵作为序列 $x = \{x_k\}$ 的代价函数[8,9],即

$$M(x) = -\sum_k p_k \log p_k$$

其中,$p_k = \dfrac{|x_k|^2}{\|x\|^2}$,并假设 $0\log 0 = 0$. 它是只满足半可加性的信息代价函数. 然而,由于 $\lambda(x) = -\sum_k |x_k|^2 \log |x_k|^2$ 是可加的,因此可以用关系式 $M(x) = \|x\|^{-2}\lambda(x) + \log\|x\|^2$ 作为代价函数,它满足在一定的均方误差条件下 $\exp M(x)$ 正比于表示信号所需的系数数目.

产生最好基的过程是首先根据式(1)和(2)把 U_j^0 分解成若干层的有限二分树. 再用小波包算法计算出 $f(x)$ 在各子空间上的系数,然后由代价函数 M 计算出各层上系数的代价函数值. 在有限二分树的关系中,把上层框称为父框,其一下层框称为子框. 最好基的算法如下:

(1)从最下层框开始,把每个代价函数值全标上“*”号.

(2)把两个子框中的值之和与其父框中的值比较,如果子框的值之和大于父框中的值,就把父框中的值标上“*”号;否则就用这个和值代替父框中的值.

(3)用新值和标“*”号的值,按步骤(2)进行到最顶层.

(4)从最高层开始,选择第一次遇到的标“*”号的框,丢弃被选择框以下的框.

这样选出的带“*”号的框,恰好对应于 U_j^0 的一组正交分解,对应于一组规范正交基,此基就是 $f(x)$ 相对于代价函数 M 的“最好基”. 对于不同类型的音频数据,用这种方法产生相应的最好基进行变换并产生索引.

3 产生音频数据索引

3.1 基于小波的索引

在基于小波索引方法中[3],把 MATLAB 中的函数 WAVDEC 用于分析音频数据. 输出分解结构包含小波分解向量 C 和系数长度向量 L. 分解向量 C 是由近似系数和详细系数组成的 DWT 系数集,它的结构按如下组:

C = [app. coef. (N) | det. coef. (N) | ⋯| det. coef. (1)]

系数长度向量 L 的各分量表示如下

L(1) = lengthof(app. coef. (N))

L(j) = lengthof(det. coef. (N - j + 2)) , j = 2, ⋯, N + 1,

其中, app. coef. (N) 表示近似系数集, det. coef. (i) 表示详细系数集且 $1 \le i \le N$, lengthof(app. coef. (N)) 近似系数集的长度, lengthof(det. coef. (N - j + 2)) 表示详细系数集的长度且 $2 \le j \le N + 1$, N 表示分解的最大级数.

从 DWT 系数集内选出合适的系数来作为索引. 在图 1 中,显示了从 DWT 系数中得出索引的两种常用的方法. 在两种方法中,所有的近似系数被选择. 在方法 1 中,分别选择不同级最前面的部分详细系数. 如,这可以 N, ⋯, $n1$ 级的详细系数的前 20 %, $n1$ ⋯$n2$ 级的详细系数的前 10 %,而 $n2$ ⋯1 级详细系数则不选择,如图 1(a). 在方法 2 中,分别选择最高值系数的 20 %,10 %等(而不是最前面的 20 %,10 %等),如图 1(b). 从 C 中通过适当系数的选择得出 C′,从 L 中反映的选择系数的数量和位置得出 L′它们组成近似和详细系数的数量(和位置), C 被用于索引而 L 作为辅助的信息. 对于方法 1 和 $N = 3$(3 级分解)的情况. 用 n_a' 和 n_i' 分别表示选择的近似和 i 级系数的数量. 那么,

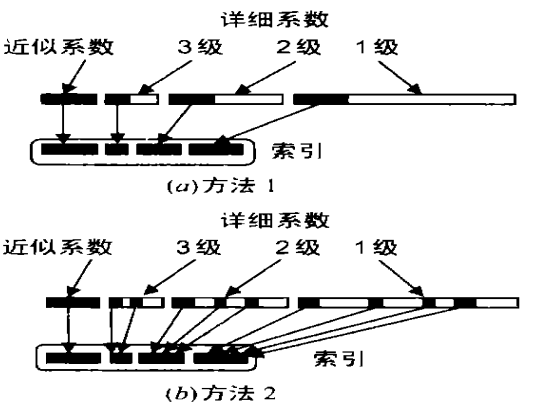

图 1 从离散小波变换中产生索引

C′= [C(1 : n_a') | C(L(1) + 1 : L(1) + n_3' - 1) | C(L(1) + L(2) + 1 : L(1) + L(2) + n_2' - 1) | C(L(1) + L(2) + L(3) + 1 : L(1) + L(2) + L(3) + n_1' - 1)]

和 $L' = [n'_a, n'_3, n'_2, n_1]$.

在方法 2 中,C 的详细系数是 C 的详细系数的不同级的最高值的某些数量的基本集合.因此,所有选择的级的详细系数可以不连续.它要求额外的定位信息存储在 L 中,如果用 $l_1, \cdots, l_{n_i}$ 表示在 L 中 i 级详细系数的位置.那么,它可以表示如下:

$$L' = [< n'_a, l_1, \cdots, l_{n'_a} > , < n'_6, l_1, \cdots, l_{n'_6} > ,$$
$$< n'_5, l_1, \cdots, l_{n'_5} > , \cdots, < n'_1, l_1, \cdots, l_{n'_1} >]$$

3.2　基于小波包最好基的索引

我们根据代价函数和最好基的产生算法,对不同类型的音频数据(如,语音、音乐和其它声音)产生不同类型的最好基.产生的最好基如图 2 所示.再用产生的最好基对待产生索引的音频数据进行小波变换.不同的最好基,通过变换,产生不同的最好基分解向量 C 和系数长度向量 L.对于图 2 所示的最好基,最好基分解向量 C 和系数长度向量 L 可以分别表示为

$$C = [coef.(000000) | coef.(000001) | coef.(00001) | coef.(0001) | \cdots | coef.(0111) | coef.(100) | coef.(101) | coef.(11)],$$
$$L(1) = lengthof(coef.(000000)),$$
$$L(2) = lengthof(coef.(000001)),$$
$$L(3) = lengthof(coef.(00001)),$$
$$L(4) = lengthof(coef.(0001)), \cdots,$$
$$L(10) = lengthof(coef.(0111)).$$

其中,$coef.(000000)$ 表示"000000"域中的系数集,$lengthof(coef.(000000))$ 表示"000000"域中系数集的长度,其余各项以此类推.

图 2　由最好基算法产生的最好基

检查变换系数向量 C 的每个系数域中系数的能量分布,在每个域中分别选取一定比例(如 5 %、10 %、20 %、40 %等)能量最大的系数作为索引.并因此产生能量子集 C 及它的辅助位置信息子集 L.对于 C 和 L 的定义与取值方式与 2.1 中"方法 2"的 C 和 L 相同.我们利用能量子集 C 及它的辅助位置信息子集 L 来作为音频数据索引的关键信息.

4　产生音频数据索引与结果分析

4.1　产生音频数据索引

当音频数据存入音频数据库时,首先对音频数据进行基于小波包最好基的变换,再分别从变换的每个域中选取能量最大的 20 %,10 %等系数的能量作为索引,并产生一组变换系数的子集及其辅助信息.把这一组信息插入到索引结构中作为原始数据的索引,并保留一个指向原始音频块的指针.当需要进行基于例子的查询时,同样把待查询的例子音频进行上述变换后,也分别从变换的每个域中选取能量最大的同样比例的系数能量作为索引,产生一组变换系数能量的子集及其辅助信息.然后,用这组信息对上述的音频索引结构进行搜索.如果两串间的相互距离足够小,它们被认为是匹配的,最后返回搜索的结果.

4.2　实验结果分析

使用 MATLAB 执行这种方法.数据文件由语音、音乐和多个其它种类的数据如动物和鸟声、钟汽笛声等特殊的声音组成.文件和查询大小大约相同(15 秒),总共 200 个音频数据文件.分别采用基于小波的索引(方法 2)和基于小波包最好基变换的索引,来实验两种方法下的回取精度.实验分成四组,每组采用的变换系数的比例分别为 5 %、10 %、20 %、40 %.每组重复 10 次,每次随机抽取 100 个数据文件.每次实验结束后,把这些数据放回,重新随机抽取作为下次实验的 100 个数据文件.实验使用 6 级 Daubechies-4 小波.两种方法匹配时,采用 Euclidean 距离来作为距离度量方式.基于小波包最好基变换的索引方法与基于小波变换的索引方法实验结果如表 1.我们使用检索精度作为指标,其定义如下:

$$\text{检索精度} = \frac{\text{返回的相关对象数}}{\text{返回的对象总数}} \times 100\ \%$$

表 1　基于小波包最好基和小波变换的索引的检索精度(%)

变换系数比例		重复实验的序数										
		1	2	3	4	5	6	7	8	9	10	平均
5 %	dwt	13	29	50	11	30	67	43	40	33	40	35.6
	wpt	17	32	60	13	43	65	45	47	56	48	42.6
10 %	dwt	14	67	67	33	67	67	40	67	43	67	53.2
	wpt	44	77	73	45	75	63	55	71	42	66	61.1
20 %	dwt	50	50	82	88	50	50	40	50	40	50	55.0
	wpt	64	79	93	93	77	73	75	69	78	68	76.9
40 %	dwt	91	90	94	89	91	50	50	93	57	50	75.5
	wpt	90	95	97	96	95	98	93	95	87	95	94.1

从表 1 可以看出,采用基于小波包最好基变换的索引的回取精度高于基于小波变换的索引,且对于采用不同比例的变换系数回取精度相对稳定.

5　结论

采用基于小波包最好基变换的索引具有较高音频数据的回取精度.与常用的基于信号统计音频数据回取方法、基于短时 Fourier 变换方法和基于小波变换的方法相比,在音频多媒体数据库中,采用这种方法无非较为理想.这种方法也为基于例子和基于内容的音频多媒体数据库查询进入实用提供了一种有效的方法.

参考文献:

[1] Wold E, Blum T, Keislar D, Wheaton J. Content-based classfication, search and retrieval of audio [J]. IEEE Multimedia, 1996, 3(3): 27-36.

[2] Chias A, Logan J, Chamberlin D, et al. Query by humming: musical information retrieval in an audio database [A]. San F. Proceedings of MULTIMEDIA'95 [C]. New York: ACM, 1995. 231-236.

[3] Subramanya S R. Indexing and Searching Schemes For Audio Data in Audio/Multimedia Databases [D]. USA: George Washington Univ., in 1999.

[4] Carnero B, Drygajlo A. Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms [J]. IEEE Trans on Signal Processing, 1999, 47(6): 1622 - 1635.

[5] Philippe P, Lever M. Wavelet packet filterbanks for low time delay audio coding [J]. IEEE Trans on Speech and Audio Processing, 1999, 7(3): 310 - 322.

[6] Kurth F, Clausen M. Filter bank tree and M-band wavelet packet algorithms In audio signal processing [J]. IEEE Transactions on Signal Processing, 1999, 47(2): 549 - 554.

[7] Wu X D, Li Y M, Chen H Y. Multi-domain speech compression based on wavelet packet transform [J]. Electronics Letters, 1998, 34(2): 154 - 155.

[8] Coifman R R, Wickerhauser M V. Entropy-Based Algorithms for Best Basis Selection [J]. IEEE Trans Information Theory, 1992, 38(2).

[9] Rao B D, Kreutz-Delgado K. An affine scaling methodology for best basis selection [J]. IEEE Transactions on Signal Processing, 1999, 47(1): 187 - 200.

作者简介:

李 应 男,1964 年 8 月生于福建闽清,1993 年在西安交通大学获系统工程专业硕士学位,现为福州大学计算机与信息科学讲师,并在西安交通大学攻读博士学位研究生,主要研究方向为计算机仿真、多媒体数据库及音频数据处理.

侯义斌 男,1952 年 4 月生于陕西武功,1982 年在西安交通大学获计算机工程与科学专业硕士学位,1986 年在 Eindhoven 技术大学获电子工程专业博士学位,现为西安交通大学计算机和信息技术教授,并为 Eindhoven 技术大学客座教授,主要研究方向包括中文信息处理、多媒体处理、用户-系统交互和设计方法学.